Using the iPhone for Voice Recording in Laryngology

Jeremy Hornibrook,[1] Emily Lin,[2] & Tika Ormond[2]

[1]Department of Otolaryngology, Christchurch Hospital
[2]Department of Communication Disorders, University of Canterbury
Christchurch, New Zealand

Canterbury District Health Board
Te Poari Hauora ō Waitaha

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Background

- iPhone (first released in: Year 2007): a multi-media enabled mobile phone with advanced
  - Computing capabilities
  - Connectivity: Internet and email access
- iPhone 3G & up: sampling rate = 48,000 Hz (Lossless)

Research Question

- Can the iPhone be used to record voices for acoustic analysis to:
  - Identify voice aberrations
  - Monitor voice changes: e.g.,
    pre- and post-treatment differences?

Participants

- Twenty-two voice patients (10 males & 12 females), aged 25-92 years (Mean = 54.8, SD = 18.5), including:
  - 10 patients (6 males & 4 females) who underwent phonosurgery:
    - Aged 33-79 years (Mean = 47.6, SD = 15.3)
    - Pathology:
      - **Mass Lesions** (8 cases): cyst (2), nodules (2), edema (1), papilloma (1), polyps (1), benign tumor (1)
        - Treatment: microsurgery
      - **Vocal Paralysis** (2 cases)
        - Treatment: medialization laryngoplasty

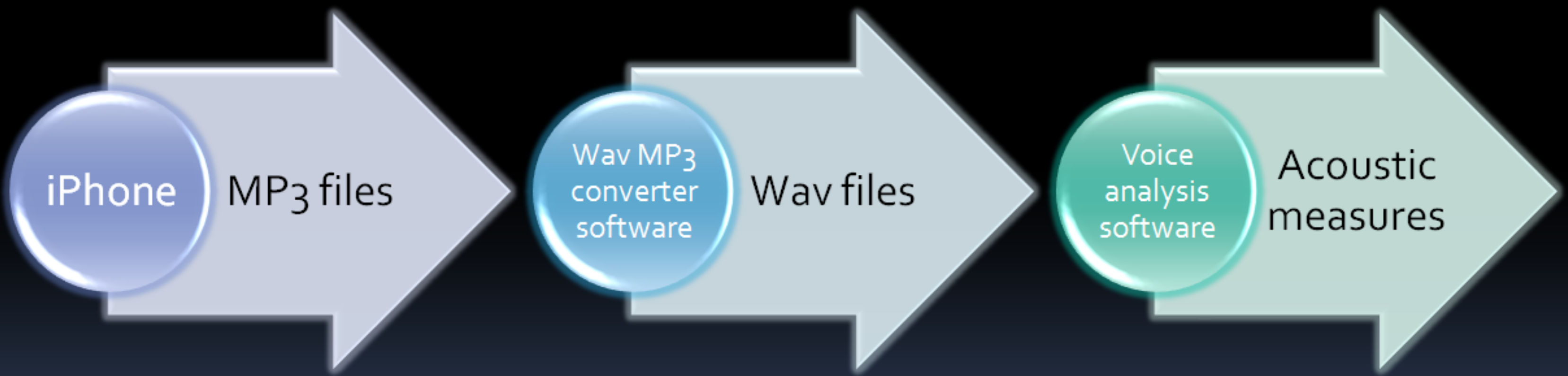
Participant's Task

- To read the first 6 sentences of "Rainbow passage" *(Fairbanks, 1960)*
- Recordings:
  - Before treatment: for all of the 22 participants
  - After treatment: for 10 of the participants
    - Time between pre- and post-surgery recordings:

      22-259 days (Mean = 103, SD = 95.5)

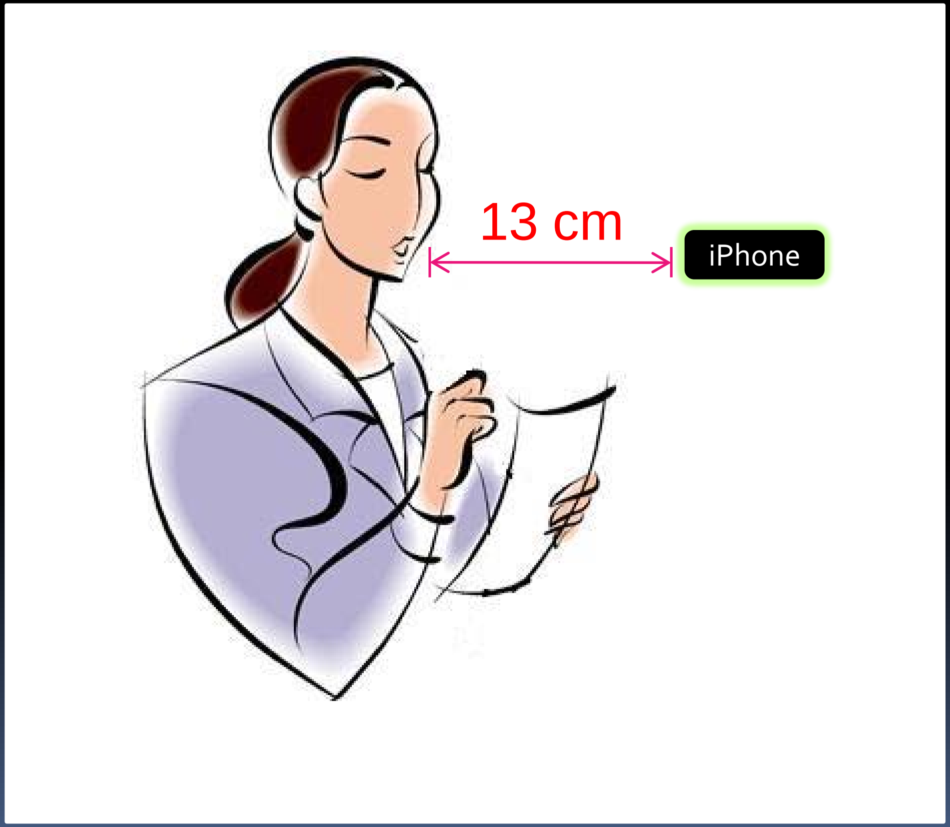
Instrumentation

- Apple iPhone (Model A1303)
- Wav MP3 converter software (Hoo Technologies, USA)
- TF32 voice analysis software *(Milenkovic, 1987)*



Procedure

- Laryngostroboscopic examination
- Acoustic recording



Case One: Left Vocal Fold Paralysis

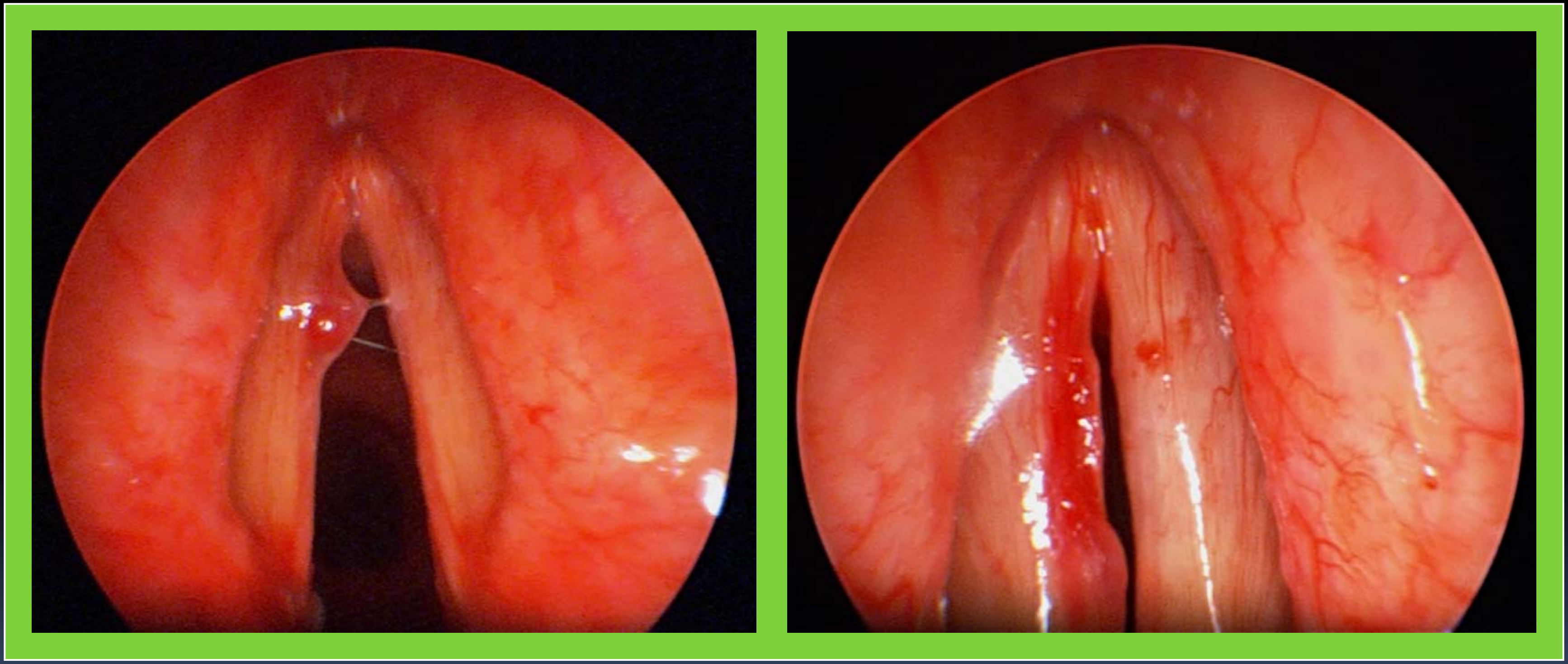
- Male, aged 79 years, aortic arch aneurysm

PRE

POST (medialization laryngoplasty)

Case Two: Polyp on LVF

- Male, aged 33 years

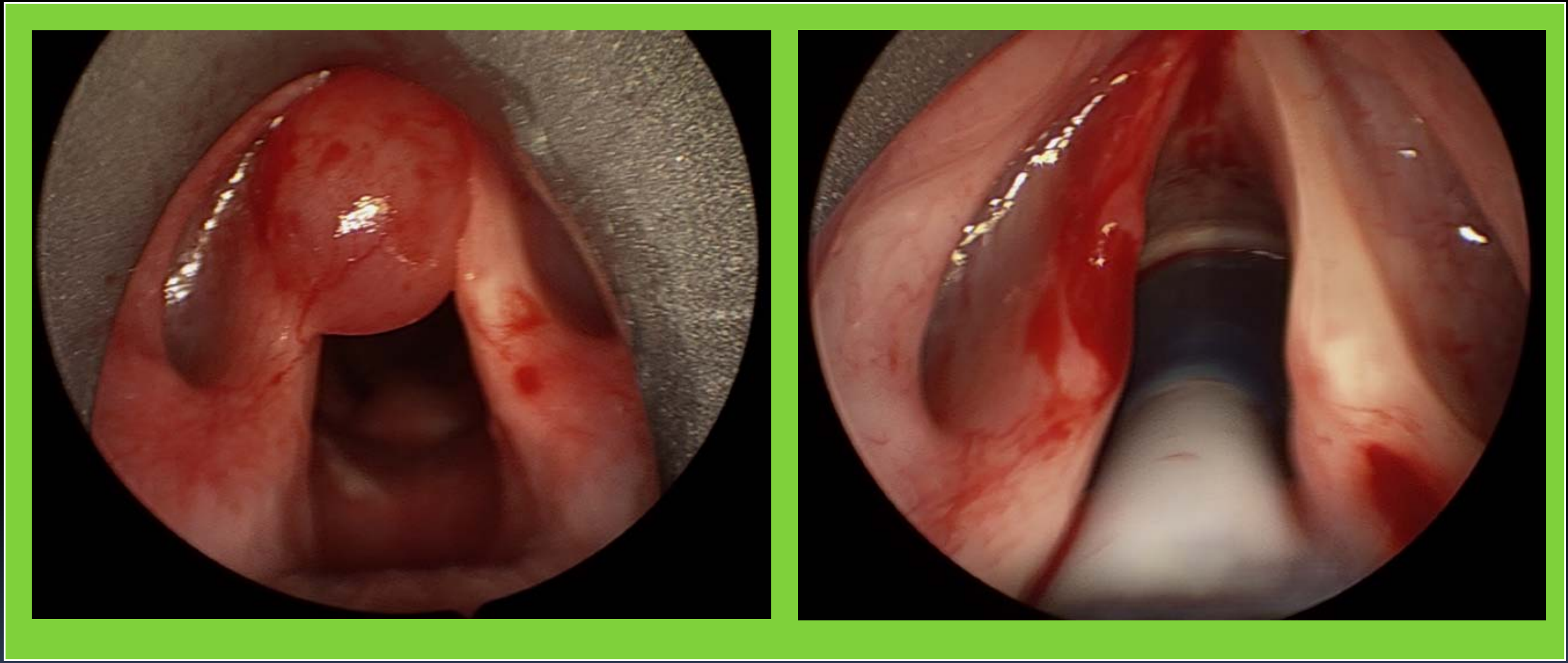


PRE

POST (microsurgery)

Case Three: Inflammatory myoblastic tumor

- Male, aged 34 years

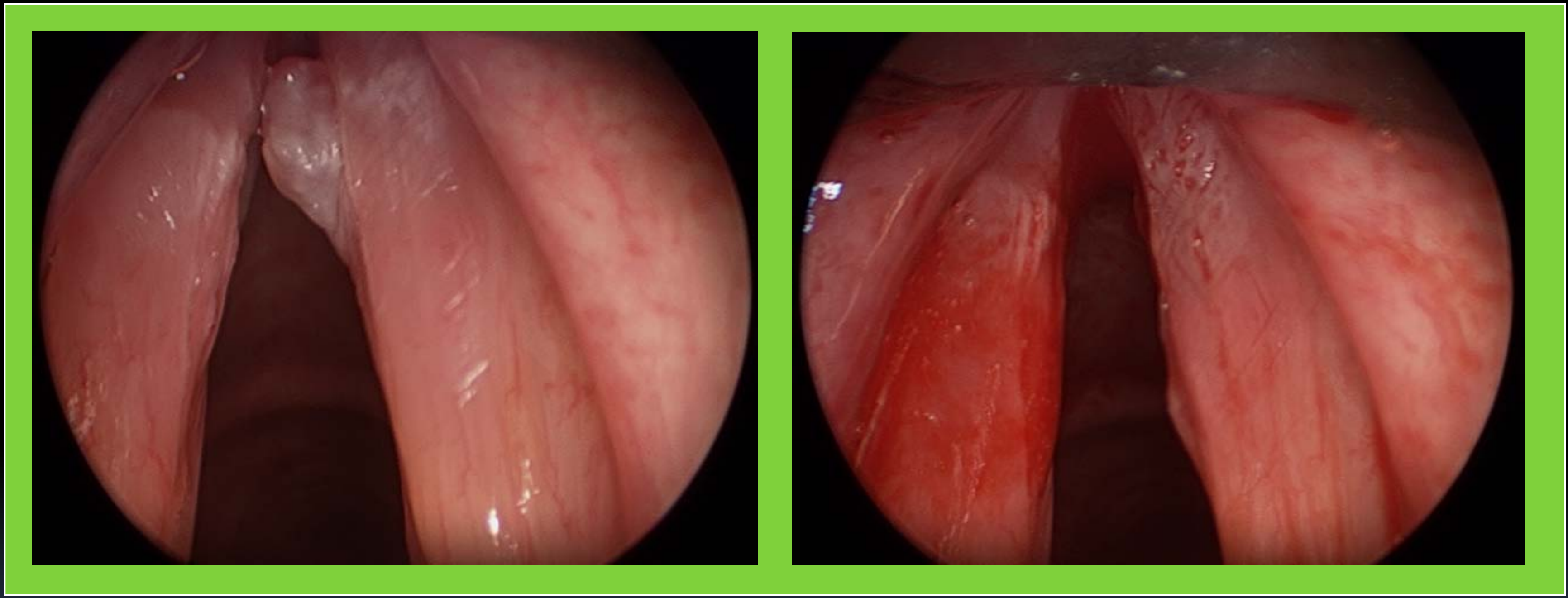


PRE

POST (microsurgery)

Case Four: Mass lesion on RVF

- Male, aged 46 years

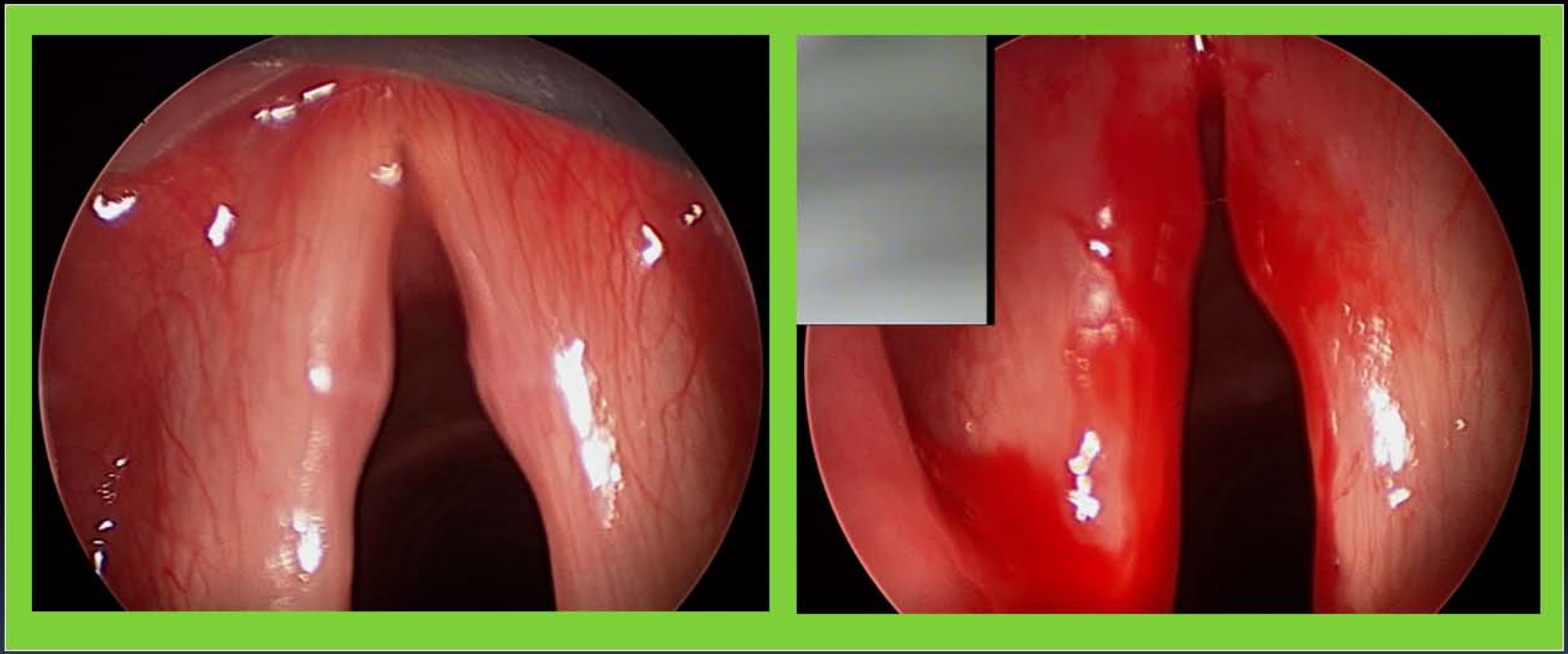


PRE

POST (microsurgery)

Case Five: Nodules

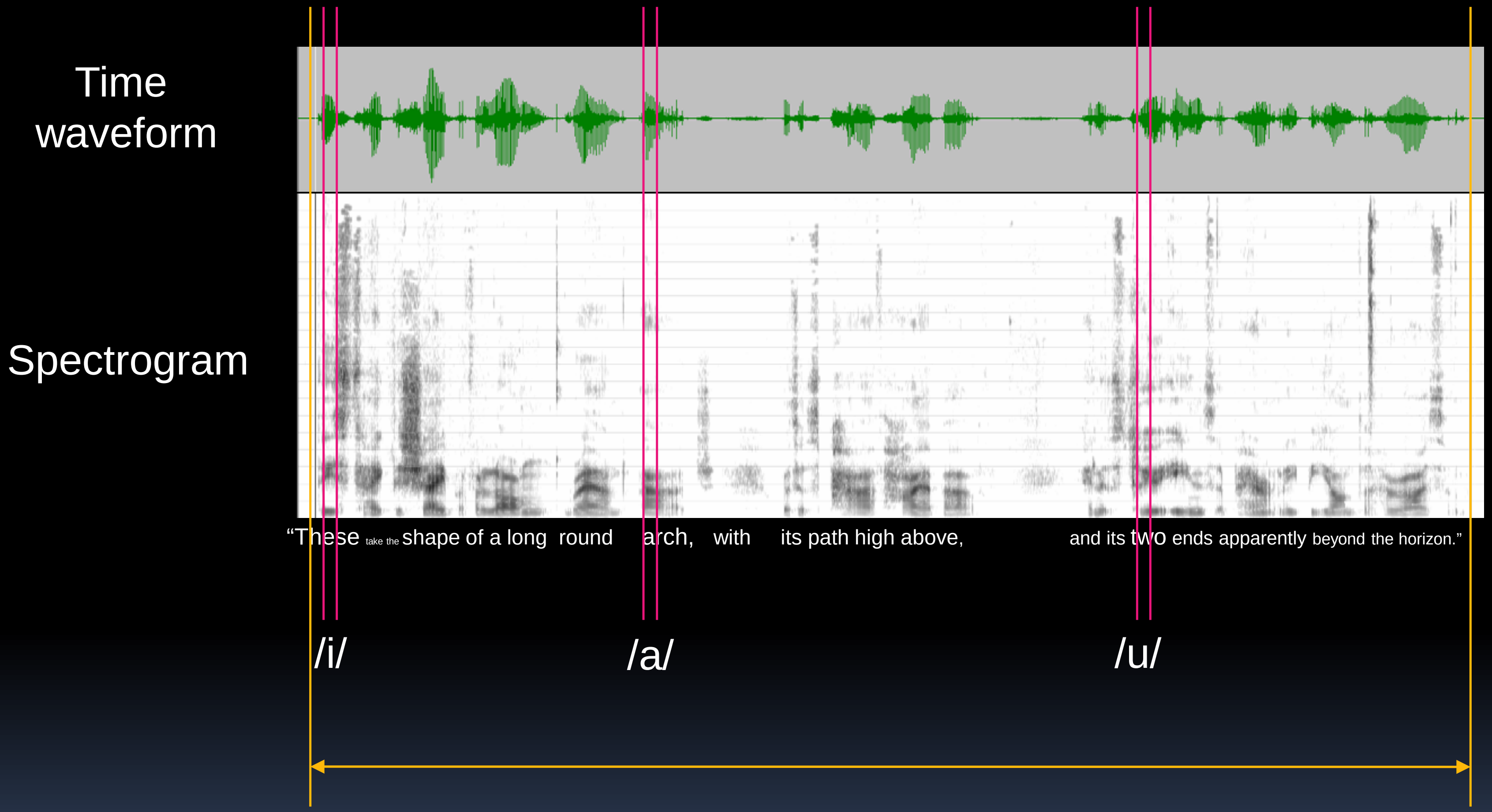
- Female, aged 39 years



PRE

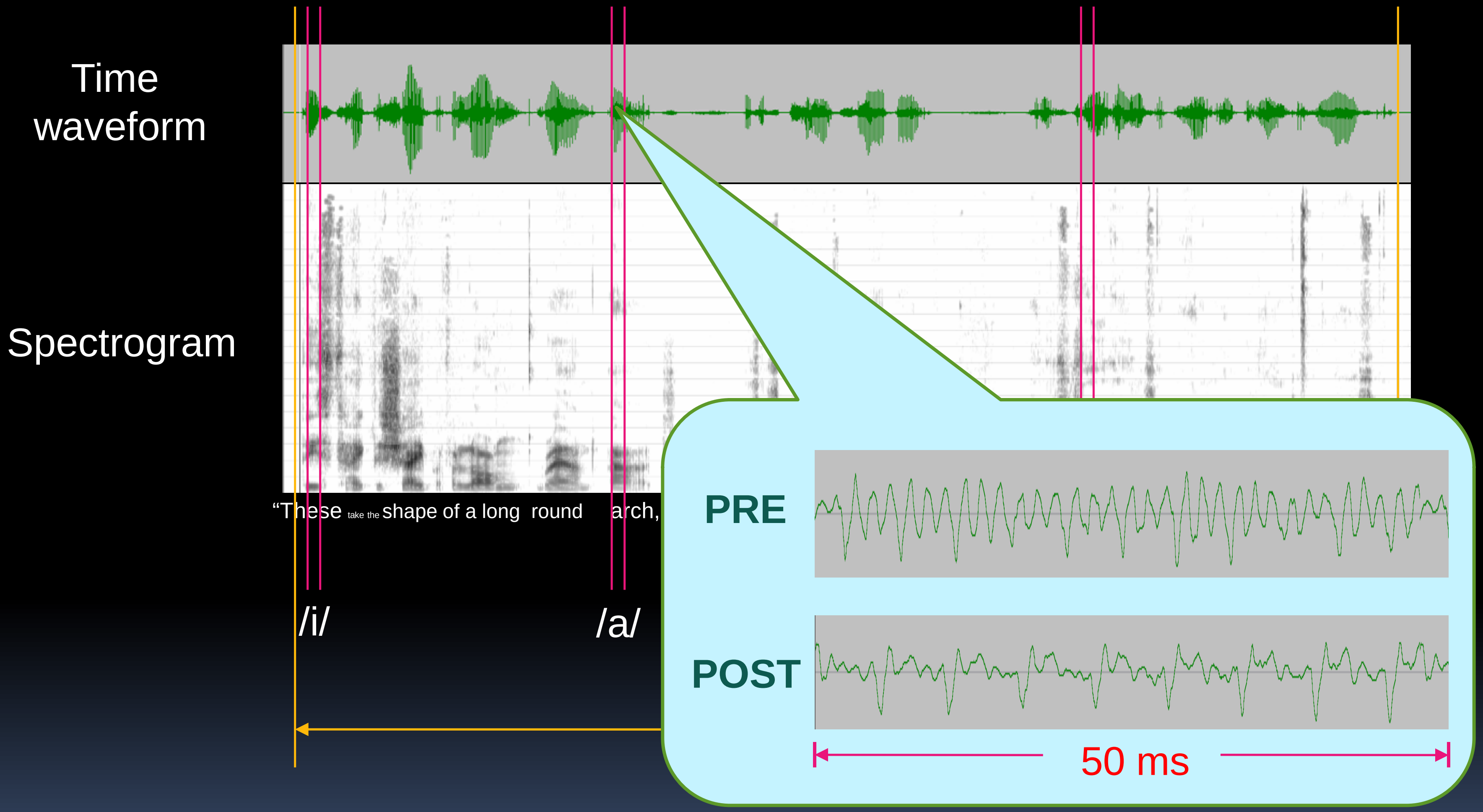
POST (microsurgery)

Data Analysis



Vowel-based measures: F0, perturbation measures (%Jit, %Shim, SNR), H1-H1, SPR, F1, F2
Sentence-based measures: ST

Data Analysis



Vowel-based measures: F0, perturbation measures (%Jit, %Shim, SNR), H1-H1, SPR, F1, F2
Sentence-based measures: ST

Measurement

- **Vowel-based measures** (time waveform analysis):
  - Fundamental frequency (F0):
    - Edema: decreased F0 *(Sorensen & Horii, 1982)*
    - Speaking F0 changes after treatment of functional voice (Roy & Taskco, 1994)

    Changes in mass and stiffness → Change in F0

  - Perturbation measures:
    - Jitter (or percent jitter; %Jit): cycle-to-cycle pitch variation *(e.g., Lieberman, 1961; Eskenazi et al., 1990; Dejonckere et al., 1996; Wolfe & Martin, 1997; Bhuta et al., 2004)*
    - Shimmer (or percent shimmer, %Shim): cycle-to-cycle amplitude variation *(e.g., Dejonckere et al., 1996; Wolfe & Martin, 1997; Bhuta et al., 2004)*
    - Signal-to-noise ratio (SNR): energy ratio between periodic & aperiodic components *(Yanagihara, 1967; Wolfe & Martin, 1997; Brockmann, Storck, Carding, & Drinnan, 2008)*

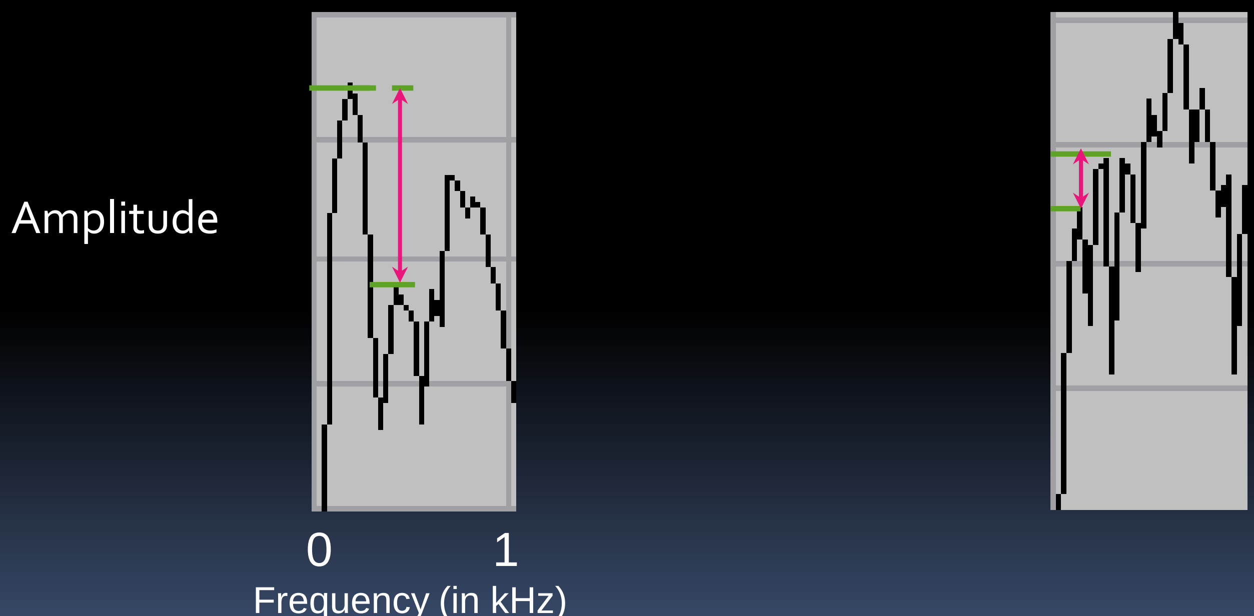
    Less hoarse → decreased jitter & shimmer; increased SNR

Measurement – *continued*

- **Vowel-based measures** (spectral analysis):
  - Amplitude difference between the first two harmonics (H1-H2) as measured from a spectrum (without pre-emphasis)

    Less breathy voice → Smaller H1-H2 (i.e. less H1 dominance)

    (Klatt & Klatt, 1990; Hillenbrand, Cleveland, & Erickson, 1994; de Krom, 1995; Hillenbrand & Houde, 1996; Stone, Cleveland, Sundberg, & Prokop, 2003)



PRE (M79-paralysis) POST-surgery

Measurement – *continued*

- **Vowel-based measures** (spectral analysis) - *continued*:
  - Singing power ratio (SPR):
    - Defined as: amplitude difference between the highest spectral peak between 0 and 2 kHz and that between 2 and 4 kHz as measured from a spectrum (with pre-emphasis)

Greater voice projection power → Smaller SPR

(Omori, Kacker, Carroll, Riley, & Blaugrund, 1996)

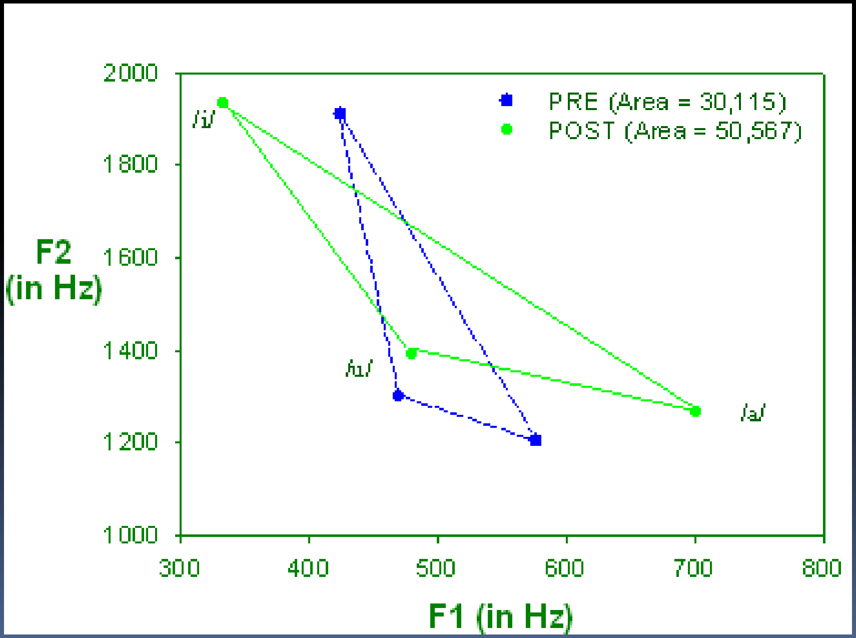
PRE (M46-paralysis)   POST-surgery

Measurement – *continued*

- **Vowel-based measures** (spectral analysis) - *continued* :
  - Formants One and Two frequencies (F1 and F2):
    - Defined as: the highest two spectral peaks in a LPC (linear predictive coding) spectrum
    - Related to vocal tract configuration (constriction or tongue placement)
    - Space enclosed by the corner vowels /i, a, u/ in a F1-F2 plot = vowel space

Increased vowel differentiation → Larger vowel space area

(Bradlow, Toretta, & Pisoni, 1996; Roy, Nissen, Dromey, & Sapir, 2009; Turner, Tjaden, & Weismer, 1995; Weismer, Jeng, Laures, Kent, & Kent, 2001)



- PRE (M34-tumor)
  - /i/
  - /a/
  - /u/
- POST-surgery
  - /i/
  - /a/
  - /u/

Measurement – continued

- **Sentence-based measures:**
  - Spectral tilt (ST): amplitude difference between the highest spectral peak between 0 and 1 kHz and that between 1 and 5 kHz as measured from a LTA (long-time average) spectrum (without pre-emphasis)

    More lax vocal fold adduction → higher ST (i.e., steeper slope)

    (Löfqvist, 1987; Mendoza et al., 1996)

PRE (F67-edema) POST-surgery

Reliability

- Automatic computer derivations of acoustic measures: 100% reliability, except for errors due to variation in vowel segmentation.
- One third of the data for the ten patients with pre- and post-treatment recordings were re-segmented and analyzed. The measure-remeasure reliabilities (Pearson's r) were high for all vowel-based measures:

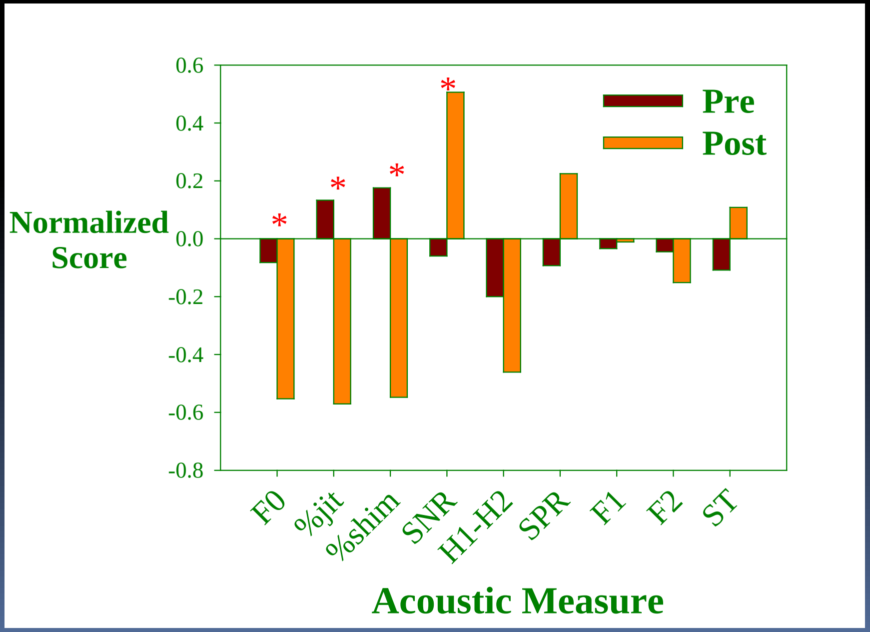
Measure	n	r
F0	20	0.99
%Jit	20	0.99
%Shim	20	0.98
SNR	20	0.97
F1	20	0.93
F2	20	0.88
H1-H2	20	0.97
SPR	20	0.96

Statistical Analysis

- A series of two-way (2 groups X 3 vowels) mixed model ANOVAs on F0, %Jit, %Shim, SNR, H1-H2, SPR, F1, & F2.
  - Between-subject variable: group (Pre vs. Post)
  - Within-subject variable: vowel (/i/, /a/, & /u/)
- A series of paired t tests on vowel space area & ST (averaged from six sentences).
- Significance level set at 0.05.

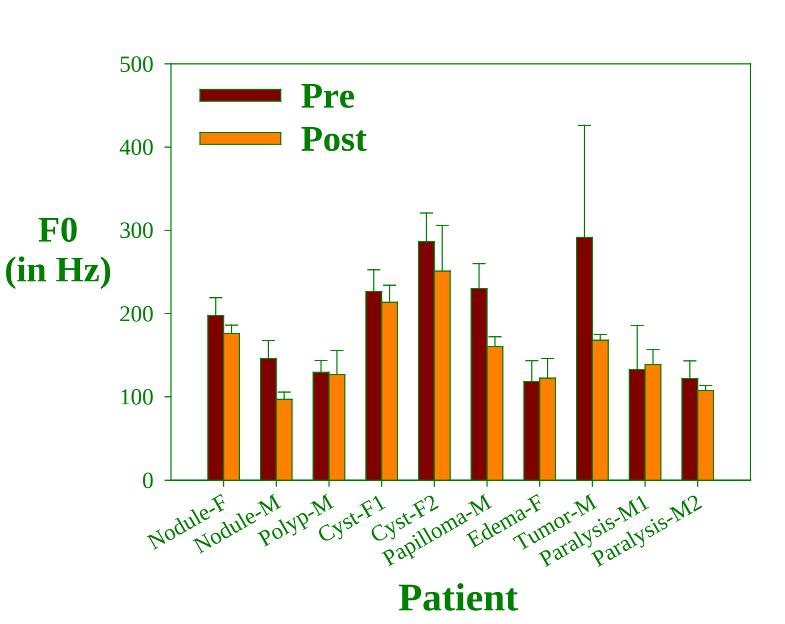
Results

- After surgery: Significantly lower F0, %Jit, & %Shim and higher SNR & vowel space area.
  - Vowel space area: significant larger vowel space area (t = -3.746, df = 9, p = 0.007)
  - ANOVA results:
    - No significant vowel by group interaction effect.
    - Significant vowel effect on F0, SNR, SPR, F1, & F2
    - Significant group effect on F0, %Jit, %Shim, & SNR.
      - Shown below with the average pre- and post-surgery scores normalized based on the pre-treatment data of 22 patients:



Results - *continued*

- F0 & H1-H2 change varied by pathology:
  - Mass lesion (8 cases): After surgery,
    - Lower pitch (except for edema): Significantly lower F0 as a whole [$F(1, 14) = 6.788$, $p = 0.035$, $\eta^2 = 0.08$].
    - No significant H1-H2 change
  - Paralysis (2 cases): After surgery,
    - No significant F0 change
    - Less breathy: Significantly lower H1-H2 [$F(1, 2) = 996.755$, $p = 0.002$, $\eta^2 = 0.71$]



Results *- continued*

- SPR change varied by vowel:
  - With the vowel /a/, the majority (7 out of 10) showed a high preoperative SPR value (i.e., less voice projection power)
  - More voice projection power: This subgroup showed a significantly lower SPR after surgery ($t = 3.383$, $df = 6$, $p = 0.015$).
- ST change varied by gender: After surgery,
  - Females ($n = 4$):
    - More lax vocal fold adduction: Significantly higher ($t = -7.683$, $df = 3$, $p = 0.005$).
  - Males ($n = 6$): no significant change

Discussion *- Continued*

- Signs of voice improvement:
  - The decrease in %Jit & %Shim and the increase in SNR found in the postoperative voices were expected as previous studies have shown that phonatory stability could be compromised by vocal pathology and improved with effective treatment.
  - The expansion of vowel space area after surgery reflects improved speech clarity, suggesting that voice quality may affect vowel intelligibility.

- No consistent F0 change after surgery for patients with vocal fold paralysis: this agrees with previous findings (LaBlance & Maves, 1992).

Limitations

- Small sample size
- Observations made in the subgroups regarding the confounding effects of pathology, vowel, and gender on the pre- and post-surgery acoustic changes require further investigations

Conclusions

- Voice recordings using iPhone are adequate for voice recording in acoustic assessment of voice quality in .
- However, due to large inter-subject variations, most of these measures are more useful for intra-subject comparison (to monitor changes within individuals) than for norm-referenced comparisons.

References

- Bhuta, T., Patrick, L., & Garnett, J. D. (2004). Perceptual evaluation of voice quality and its correlation with acoustic measurements. *Journal of Voice, 18,* 299-304.
- Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (1996). Intelligibility of normal speech I: global and fine grained acoustic-phonetic talker characteristics. *Speech Communication, 20,* 255-272.
- Brockmann, M., Storck, C., Carding, P. N., & Drinnan, M. J. (2008). Voice loudness and gender effect on jitter and shimmer in healthy adults. *Journal of Speech, Language, and Hearing Research*, *51,* 1152-1160.
- de Krom, G. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *Journal of Speech and Hearing Research, 38,* 794-811.
- Dejonckere, P. H., Remacle, M., Fresnel-Ebaz, E., Woisard, V., Crevier-Buchman, L., & Millet, B. (1996). Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements. Rev Laryngol Otol Rhinol (Bord), 117, 219-224.
- Eskenazi, L., Childers, D. G., & Hicks, D. (1990). Acoustic correlates of vocal quality. *Journal of Speech and Hearing Research, 33,* 298-306.
- Fairbanks, G. (1960). Voice and Articulation Drillbook. 2nd ed. New York: Harper & Row. Pg. 127.

References - continued

- Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy voice quality. *Journal of Speech and Hearing Research, 37,* 769-778.
- Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech and Hearing Research, 39,* 311-321.
- Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87,* 820-854.
- Kotby, M. N., Titze, I. R., Saleh, M. M., & Berry, D. A. (1993). Fundamental frequency stability in functional dysphonia. *Acta Otolaryngologica, 113,* 4439-444.
- LaBlance G. R., & Maves, M. D. (1992). Acoustic characteristics of post-thyroplasty patients. *Otolarygnol Head Neck Surg, 107,* 558-563.
- Löfqvist, A., & Mandersson, B. (1987). Long-time average spectrum of speech and voice analysis. *Folia Phoniatrica (Basel), 21,* 221-229.
- Mendoza, E., Munoz, J., Valencia Naranjo, N. (1996). The long-term average spectrum as a measure of voice stability. *Folia Phoniatrica et* Logopaedica, *48,* 57–64.

References - continued

- Omori, K., Kacker, A., Carroll, L. M., Reiley, W. D., & Blaugrund, S. M. (1996). Singing power ratio: quantitative evaluation of singing voice quality. *Journal of Voice, 10,* 228-235.
- Roy, N., Nissen, S. L., Dromey, C., & Sapir, S. (2009). Articulatory changes in muscle tension dysphonia: evidence of vowel space expansion following manual circumlagyngeal therapy. *Journal of Communication Disorders, 42,* 124-135.
- Roy, N., & Tasko, S. M. (1994). Speaking fundamental frequency (SFF) changes following successful management of functional dysphonia. Journal of Speech-Language Pathology and Audiology, 18, 115-120.
- Sorensen, D., & Horii, Y. (1982). Cigarette smoking and voice fundamental frequency. *Journal of Communication Disorders, 15,* 135-144.
- Stone, R. E. Jr., Cleveland, T. F., Sundberg, P. J., & Prokop, J. (2003). Aerodynamic and acoustical measures of speech, operative, and Broadway vocal styles in a professional female singer. *Journal of Voice, 17,* 283-297.

References - continued

- Turner, G. S., Tjaden, K., & Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amytrophic lateral sclerosis. *Journal of Speech and Hearing Research, 38,* 1001-1013.
- Watts, C. Barnes-Burroughs, K., Estis, J., Blanton, D. (2006). The singing power ratio as an objective measure of singing voice quality in untrained talented and nontalented singers. *J. of Voice, 20,* 82-88.
- Weismer, G., Jeng, J-Y., Laures, J. S., Kent, R. D., & Kent, J. F. (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatrica, 53,* 1-18.
- Winholtz, W. S., & Titze, I. R. (1998). Suitability of minidisc (MD) recordings for voice perturbation analysis. *J. of Voice, 12,* 138-142.
- Wolfe, V. & Martin, D. (1997). Acoustic correlates of dysphonia: type and severity. *Journal of Communication Disorders, 30,* 403-416.
- Yanagihara, N. (1967). Significance of harmonic chagnes and noise components in hoarseness. *J Speech Hear Res, 10,* 531-541.